Express Mail Label No.
EL844512138US
Date of Deposit: 07-31-01

PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER

Related Applications

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. provisional application 60/221,735, filed July 31, 2000.

Field of the Invention

The invention relates to nucleic acid microarray markers for cancer, particularly for endometrial cancer. The invention also relates to methods for diagnosing cancer as well as optimizing cancer treatment strategies.

Background of the Invention

Endometrioid endometrial adenocarcinomas are the most common gynecologic malignancy, the risk of which is increased by an abnormal endocrine environment or premalignant lesions with loss of tumor suppressor function. The 6000 deaths yearly make uterine cancer the seventh leading cause of death from malignancy in females. It is primarily a disease of postmenopausal women, although 25 percent of cases occur in women below age 50 and 5 percent below age 40 (Harrison's Principles of Internal Medicine 1998).

Although much progress has been made toward understanding the biological basis of cancer and in its diagnosis and treatment, it is still one of the leading causes of death in the United States. Inherent difficulties in the diagnosis and treatment of cancer include among other things, the existence of many different subgroups of cancer and the concomitant variation in appropriate treatment strategies to maximize the likelihood of positive patient outcome.

The prognosis of endometrial cancer depends upon stage, histologic grade, and extent of myometrial invasion. The staging of endometrial cancer requires surgery to establish the extent of disease and the depth of myometrial invasion. Peritoneal fluid should be sampled; the abdomen and pelvis explored; and pelvic and para-aortic lymphadenectomy performed depending upon the histology, grade, and depth of invasion in the uterine specimen on frozen section.

Initial evaluation of patients suspected of endometrial cancer includes a history and physical and pelvic examination followed by an endometrial biopsy or a fractional dilation

and curettage. Outpatient procedures such as endometrial biopsy or aspiration curettage can be used but are definitive only when positive. Once a diagnosis is made, the options for treating endometrial cancer are assessed with respect to the needs of the patient. These options traditionally include surgical intervention, radiotherapy, chemotherapy, and adjuvant systemic therapies. Adjuvants may include but are not limited to chemotherapy, radiotherapy, and endocrine therapies with progestational agents such as hydroxyprogesterone, megastrol, and deoxyprogesterone, and the antiestrogen tamoxifen.

It is difficult to predict from standard clinical and pathologic features the clinical course of endometrial cancer. However, it is very important in the treatment of endometrial cancer to select and implement an appropriate combination of therapeutic approaches. The available methods for designing strategies for treating endometrial cancer patients are complex and time consuming. The wide range of cancer subgroups and variations in disease progression limit the predictive ability of the healthcare professional. In addition, continuing development of novel treatment strategies and therapeutics will result in the addition of more variables to the already complex decision-making process involving matching the cancer patient with a treatment regimen that is appropriate and optimized for the cancer stage, extent of myometrial invasion, tumor growth rate, and other factors central to the individual patient's prognosis. Because of the critical importance of selecting appropriate treatment regimens for endometrial cancer patients, the development of guidelines for treatment selection is of key interest to those in the medical community and their patients. Thus, there presently is a need for objective, reproducible, and sensitive methods for predicting endometrial cancer patient outcome and selecting optimal treatment regimens.

Summary of the Invention

It now has been discovered that particular sets of genes are expressed differentially in normal and malignant endometrium. These sets of genes can be used to discriminate between normal and malignant endometrial tissues. Accordingly, diagnostic assays for classification of tumors, prediction of tumor outcome, selecting and monitoring treatment regimens, and monitoring tumor progression/regression can now be based on the expression of sets of genes.

According to one aspect of the invention, methods for diagnosing endometrial cancer in a subject suspected of having endometrial cancer are provided. The methods include

obtaining from the subject an endometrial tissue sample and determining the expression of a set of nucleic acid molecules or expression products thereof in the endometrial tissue sample. The set of nucleic acid molecules includes at least two nucleic acid molecules selected from the group consisting of SEQ ID NOs:1-50. In preferred embodiments, the endometrial tissue sample is suspected of being cancerous.

In some embodiments the set of nucleic acid molecules includes more than 2, and up to all of the nucleic acid molecules set forth as SEQ ID NOs:1-50, and any number of nucleic acid sequences between these two numbers. For example, in certain embodiments the set includes at least 3, 4, 5, 10, 15, 20, 30, 40 or more nucleic acid molecules of the nucleic acid molecules set forth as SEQ ID NOs:1-50.

In other embodiments, the method further includes determining the expression of the set of nucleic acid molecules or expression products thereof in a non-cancerous endometrial tissue sample, and comparing the expression of the set of nucleic acid molecules or expression products thereof in the endometrial tissue sample suspected of being cancerous and the non-cancerous endometrial tissue sample.

The invention in another aspect provides solid-phase nucleic acid molecule arrays. The arrays have a cancer gene marker set that consists essentially of at least two and as many as all of the nucleic acid molecules set forth as SEQ ID NOs:1-50 fixed to a solid substrate. The set of nucleic acid markers can include any number of nucleic acid sequences between these two numbers, selected from SEQ ID NOs:1-50. For example, in certain embodiments the set includes at least 3, 4, 5, 10, 15, 20, 30, 40 or more nucleic acid molecules of the nucleic acid molecules set forth as SEQ ID NOs:1-50. In some embodiments, the solid-phase nucleic acid molecule array also includes at least one control nucleic acid molecule.

In certain embodiments, the solid substrate includes a material selected from the group consisting of glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. Preferably the substrate is glass.

In other embodiments, the nucleic acid molecules are fixed to the solid substrate by covalent bonding.

According to yet another aspect of the invention, protein microarrays are provided. The protein microarrays include antibodies or antigen-binding fragments thereof, that specifically bind at least two different polypeptides selected from the group consisting of SEQ ID NOs:51-100, fixed to a solid substrate. In some embodiments, the microarray

comprises antibodies or antigen-binding fragments thereof, that bind specifically to least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 different polypeptides selected from the group consisting of SEQ ID NOs:51-100. In certain embodiments, the microarray also includes an antibody or antigen-binding fragment thereof, that binds specifically to a cancer-associated polypeptide other than those selected from the group consisting of SEQ ID NOs:51-100, preferably An endometrial cancer associated polypeptide. In some embodiments, the protein microarray also includes at least one control polypeptide molecule. In further embodiments, the antibodies are monoclonal or polyclonal antibodies. In other embodiments, the antibodies are chimeric, human, or humanized antibodies. In some embodiments, the antibodies are single chain antibodies. In still other embodiments, the antigen-binding fragments are $F(ab')_2$, Fab, Fd, or Fv fragments.

According to yet another aspect of the invention, protein microarrays are provided. The protein microarrays include antibodies or antigen-binding fragments thereof, that specifically bind at least two different polypeptides selected from the group consisting of SEQ ID NOs:51-100, fixed to a solid substrate. In some embodiments, the microarray comprises antibodies or antigen-binding fragments thereof, that bind specifically to least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 different polypeptides selected from the group consisting of SEQ ID NOs:51-100. In certain embodiments, the microarray also includes an antibody or antigen-binding fragment thereof, that binds specifically to a cancer-associated polypeptide other than those selected from the group consisting of SEQ ID NOs:51-100, preferably an endometrial cancer associated polypeptide. In some embodiments, the protein microarray also includes at least one control polypeptide molecule. In further embodiments, the antibodies are monoclonal or polyclonal antibodies. In other embodiments, the antibodies are chimeric, human, or humanized antibodies. In some embodiments, the antibodies are single chain antibodies. In still other embodiments, the antigen-binding fragments are $F(ab')_2$, Fab, Fd, or Fv fragments.

In a further aspect of the invention, methods for identifying lead compounds for a pharmacological agent useful in the treatment of endometrial cancer are provided. The methods include contacting an endometrial cancer cell or tissue with a candidate pharmacological agent, and determining the expression of a set of nucleic acid molecules in

the endometrial cancer cell or tissue sample under conditions which, in the absence of the candidate pharmacological agent, permit a first amount of expression of the set of nucleic acid molecules. The set of nucleic acid molecules includes at least two and as many as all of the nucleic acid molecules set forth as SEQ ID NOs:1-50. The methods also include detecting a test amount of the expression of the set of nucleic acid molecules, wherein a decrease in the test amount of expression in the presence of the candidate pharmacological agent relative to the first amount of expression indicates that the candidate pharmacological agent is a lead compound for a pharmacological agent that is useful in the treatment of endometrial cancer.

In some embodiments of any of the foregoing methods and products, the differences in the expression of the nucleic acid molecules are determined by nucleic acid hybridization or nucleic acid amplification methods. Preferably the nucleic acid hybridization is performed using a solid-phase nucleic acid molecule array. In other embodiments, the differences in the expression of the nucleic acid molecules are determined by protein expression analysis, preferably SELDI mass spectroscopy.

These and other aspects of the invention will be described in greater detail below.

Detailed Description of the Invention

The invention described herein relates to the identification of a set of genes expressed in endometrial cancer tissue that are predictive of the clinical outcome of the cancer. Changes in cell phenotype in cancer are often the result of one or more changes in the genome expression of the cell. Some genes are expressed in tumor cells, and not in normal cells. In addition, different genes are expressed in different subgroups of endometrial cancers, which have different prognoses and require different treatment regimens to optimize patient outcome. The differential expression of endometrial cancer genes can be examined by the assessment of nucleic acid or protein expression in the endometrial cancer tissue.

The genes identified permit, *inter alia*, rapid screening of cancer samples by nucleic acid microarray hybridization or protein expression technology to determine the expression of the specific genes and thereby to predict the outcome of the cancer. Such screening is beneficial, for example, in selecting the course of treatment to provide to the cancer patient, and to monitor the efficacy of a treatment.

The invention differs from traditional endometrial cancer diagnostic and classification techniques with respect to the speed, simplicity, and reproducibility of the cancer diagnostic assay. The invention also presents targets for drug development because it identifies genes that are differentially expressed in outcome endometrial tumors, which can be utilized in the development of drugs to treat such tumors, e.g., by reducing expression of the genes or reducing activity of proteins encoded by the genes.

The invention simplifies prognosis determination by providing an identified set of genes whose expression in endometrial cancers predicts clinical outcome as defined by tumor metastasis, recurrence, or death. In the invention RNA expression phenotyping was performed using high density microarrays generated from quantitative expression data on over 5000 (estimated 5800) genes, which have been analyzed to identify 50 specific probe sets (genes). The expression gene set has multifold uses including, but not limited to, the following examples. The expression gene set may be used as a prognostic tool for endometrial cancer patients, to make possible more finely tuned diagnosis of endometrial cancer and allow healthcare professionals to tailor treatment to individual patients' needs. The invention can also assess the efficacy of endometrial cancer treatment by determining progression or regression of endometrial cancer in patients before, during, and after endometrial cancer treatment. Another utility of the expression gene set is in the biotechnology and pharmaceutical industries' research on disease pathway discovery for therapeutic targeting. The invention can identify alterations in gene expression in endometrial cancer and can also be used to uncover and test candidate pharmaceutical agents to treat endometrial cancer.

As used herein, a subject is a human, non-human primate, cow, horse, pig, sheep, goat, dog, cat, or rodent. In all embodiments human subjects are preferred. Preferably the subject is a human either suspected of having endometrial cancer, or having been diagnosed with endometrial cancer. In a preferred embodiment of the invention the cancer is endometroid endometrial adenocarcinoma. Methods for identifying subjects suspected of having endometrial cancer may include physical examination, subject's family medical history, subject's medical history, endometrial biopsy, or a number of imaging technologies such as ultrasonography, computed tomography, magnetic resonance imaging, magnetic resonance spectroscopy, or positron emission tomography. Diagnostic methods for

endometrial cancer and the clinical delineation of endometrial cancer diagnoses are well known to those of skill in the medical arts.

As used herein, endometrial tissue sample is tissue obtained from an endometrial tissue biopsy using methods well known to those of ordinary skill in the related medical arts. The phrase "suspected of being cancerous" as used herein means an endometrial cancer tissue sample believed by one of ordinary skill in the medical arts to contain cancerous cells. Methods for obtaining the sample from the biopsy include gross apportioning of a mass, microdissection, laser-based microdissection, cytologic sampling of the endometrium using a brush, aspiration curettage, fractional dilation and curettage, or other art-known cell-separation methods.

Because of the variability of the cell types in diseased-tissue biopsy material, and the variability in sensitivity of the diagnostic methods used, the sample size required for analysis may range from 1, 10, 50, 100, 200, 300, 500, 1000, 5000, 10,000, to 50,000 or more cells. The appropriate sample size may be determined based on the cellular composition and condition of the biopsy and the standard preparative steps for this determination and subsequent isolation of the nucleic acid for use in the invention are well known to one of ordinary skill in the art. An example of this, although not intended to be limiting, is that in some instances a sample from the biopsy may be sufficient for assessment of RNA expression without amplification, but in other instances the lack of suitable cells in a small biopsy region may require use of RNA conversion and/or amplification methods or other methods to enhance resolution of the nucleic acid molecules. Such methods, which allow use of limited biopsy materials, are well known to those of ordinary skill in the art and include, but are not limited to: direct RNA amplification, reverse transcription of RNA to cDNA, amplification of cDNA, or the generation of radio-labeled nucleic acids.

As used herein, the phrase "determining the expression of a set of nucleic acid molecules in the endometrial tissue" means identifying RNA transcripts in the tissue sample by analysis of nucleic acid or protein expression in the tissue sample. As used herein, "set" refers to a group of nucleic acid molecules that include 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50 different nucleic acid sequences from the group of nucleic acid sequences numbered 1 through 50 in Table 1 (SEQ ID NOs: 1-50).

The expression of the set of nucleic acid molecules in the sample from the endometrial cancer patient can be compared to the expression of the set of nucleic acid molecules in a sample of endometrial tissue that is non-cancerous. As used herein, non-cancerous endometrial tissue means tissue determined by one of ordinary skill in the medical art to have no evidence of endometrial cancer based on standard diagnostic methods including, but not limited to, histologic staining and microscopic analysis.

Nucleic acid markers for cancer are nucleic acid molecules that by their presence or absence indicate the presence of absence of endometrial cancer. In tissue, certain nucleic acid molecules are expressed at different levels depending on whether tissue is non-cancerous or cancerous.

Hybridization methods for nucleic acids are well known to those of ordinary skill in the art (see, e.g. *Molecular Cloning: A Laboratory Manual*, J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1989, or *Current Protocols in Molecular Biology*, F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York). The nucleic acid molecules from an endometrial cancer tissue sample hybridize under stringent conditions to nucleic acid markers expressed in endometrial cancer. In one embodiment the markers are sets of two or more of the nucleic acid molecules as set forth in SEQ ID NOs: 1 through 50.

The endometrial cancer nucleic acid markers disclosed herein are known genes and fragments thereof. It may be desirable to identify variants of those genes, such as allelic variants or single nucleotide polymorphisms (SNPs) in tissues. Accordingly, methods for identifying endometrial cancer nucleic acid markers, including variants of the disclosed full-length cDNAs, genomic DNAs, and SNPs are also included in the invention. The methods include contacting a nucleic acid sample (such as a cDNA library, genomic library, genomic DNA isolate, etc.) with a nucleic acid probe or primer derived from one of SEQ ID NOs:1-50. The nucleic acid sample and the probe or primer hybridize to complementary nucleotide sequences of nucleic acids in the sample, if any are present, allowing detection of nucleic acids related to SEQ ID NOs: 1-50. Preferably the probe or primer is detectably labeled. The specific conditions, reagents, and the like can be selected by one of ordinary skill in the art to selectively identify nucleic acids related to sets of two or more of SEQ ID NOs:1 through 50. The isolated nucleic acid molecule can be sequenced according to standard procedures.

In addition to native nucleic acid markers (SEQ ID NOs:1-50), the invention also includes degenerate nucleic acids that include alternative codons to those present in the native materials. For example, serine residues are encoded by the codons TCA, AGT, TCC, TCG, TCT, and AGC. Each of the six codons is equivalent for the purposes of encoding a serine residue. Similarly, nucleotide sequence triplets that encode other amino acid residues include, but are not limited to: CCA, CCC, CCG, and CCT (proline codons); CGA, CGC, CGG, CGT, AGA, and AGG (arginine codons); ACA, ACC, ACG, and ACT (threonine codons); AAC and AAT (asparagine codons); and ATA, ATC, and ATT (isoleucine codons). Other amino acid residues may be encoded similarly by multiple nucleotide sequences. Thus, the invention embraces degenerate nucleic acids that differ from the biologically isolated nucleic acids in codon sequence due to the degeneracy of the genetic code.

The invention also provides modified nucleic acid molecules, which include additions, substitutions, and deletions of one or more nucleotides such as the allelic variants and SNPs described above. In preferred embodiments, these modified nucleic acid molecules and/or the polypeptides they encode retain at least one activity or function of the unmodified nucleic acid molecule and/or the polypeptides, such as hybridization, antibody binding, etc. In certain embodiments, the modified nucleic acid molecules encode modified polypeptides, preferably polypeptides having conservative amino acid substitutions. As used herein, a "conservative amino acid substitution" refers to an amino acid substitution which does not alter the relative charge or size characteristics of the protein in which the amino acid substitution is made. Conservative substitutions of amino acids include substitutions made amongst amino acids within the following groups: (a) M, I, L, V; (b) F, Y, W; (c) K, R, H; (d) A, G; (e) S, T; (f) Q, N; and (g) E, D. The modified nucleic acid molecules are structurally related to the unmodified nucleic acid molecules and in preferred embodiments are sufficiently structurally related to the unmodified nucleic acid molecules so that the modified and unmodified nucleic acid molecules hybridize under stringent conditions known to one of skill in the art.

For example, modified nucleic acid molecules that encode polypeptides having single amino acid changes can be prepared for use in the methods and products disclosed herein. Each of these nucleic acid molecules can have one, two, or three nucleotide substitutions exclusive of nucleotide changes corresponding to the degeneracy of the genetic code as described herein. Likewise, modified nucleic acid molecules that encode polypeptides

having two amino acid changes can be prepared, which have, e.g., 2-6 nucleotide changes. Numerous modified nucleic acid molecules like these will be readily envisioned by one of skill in the art, including for example, substitutions of nucleotides in codons encoding amino acids 2 and 3, 2 and 4, 2 and 5, 2 and 6, and so on. In the foregoing example, each combination of two amino acids is included in the set of modified nucleic acid molecules, as well as all nucleotide substitutions that code for the amino acid substitutions. Additional nucleic acid molecules that encode polypeptides having additional substitutions (i.e., 3 or more), additions or deletions [e.g., by introduction of a stop codon or a splice site(s)] also can be prepared and are embraced by the invention as readily envisioned by one of ordinary skill in the art. Any of the foregoing nucleic acids can be tested by routine experimentation for retention of structural relation to or activity similar to the nucleic acids disclosed herein.

In the invention, standard hybridization techniques of microarray technology are utilized to assess patterns of nucleic acid expression and identify nucleic acid marker expression. Microarray technology, which is also known by other names including: DNA chip technology, gene chip technology, and solid-phase nucleic acid array technology, is well known to those of ordinary skill in the art and is based on, but not limited to, obtaining an array of identified nucleic acid probes on a fixed substrate, labeling target molecules with reporter molecules (e.g., radioactive, chemiluminescent, or fluorescent tags such as fluorescein, Cye3-dUTP, or Cye5-dUTP), hybridizing target nucleic acids to the probes, and evaluating target-probe hybridization. A probe with a nucleic acid sequence that perfectly matches the target sequence will, in general, result in detection of a stronger reporter-molecule signal than will probes with less perfect matches. Many components and techniques utilized in nucleic acid microarray technology are presented in *The Chipping Forecast*, Nature Genetics, Vol.21, Jan 1999, the entire contents of which is incorporated by reference herein.

According to the present invention, microarray substrates may include but are not limited to glass, silica, aluminosilicates, borosilicates, metal oxides such as alumina and nickel oxide, various clays, nitrocellulose, or nylon. In all embodiments a glass substrate is preferred. According to the invention, probes are selected from the group of nucleic acids including, but not limited to: DNA, genomic DNA, cDNA, and oligonucleotides; and may be natural or synthetic. Oligonucleotide probes preferably are 20 to 25-mer oligonucleotides and DNA/cDNA probes preferably are 500 to 5000 bases in length, although other lengths

may be used. Appropriate probe length may be determined by one of ordinary skill in the art by following art-known procedures. In one embodiment, preferred probes are sets of two or more of the nucleic acid molecules set forth as SEQ ID NO: 1 through 50 (see also Table 1). Probes may be purified to remove contaminants using standard methods known to those of ordinary skill in the art such as gel filtration or precipitation.

In one embodiment, the microarray substrate may be coated with a compound to enhance synthesis of the probe on the substrate. Such compounds include, but are not limited to, oligoethylene glycols. In another embodiment, coupling agents or groups on the substrate can be used to covalently link the first nucleotide or olignucleotide to the substrate. These agents or groups may include, but are not limited to: amino, hydroxy, bromo, and carboxy groups. These reactive groups are preferably attached to the substrate through a hydrocarbyl radical such as an alkylene or phenylene divalent radical, one valence position occupied by the chain bonding and the remaining attached to the reactive groups. These hydrocarbyl groups may contain up to about ten carbon atoms, preferably up to about six carbon atoms. Alkylene radicals are usually preferred containing two to four carbon atoms in the principal chain. These and additional details of the process are disclosed, for example, in U.S. Patent 4,458,066, which is incorporated by reference in its entirety.

In one embodiment, probes are synthesized directly on the substrate in a predetermined grid pattern using methods such as light-directed chemical synthesis, photochemical deprotection, or delivery of nucleotide precursors to the substrate and subsequent probe production.

In another embodiment, the substrate may be coated with a compound to enhance binding of the probe to the substrate. Such compounds include, but are not limited to: polylysine, amino silanes, amino-reactive silanes (Chipping Forecast, 1999) or chromium (Gwynne and Page, 2000). In this embodiment, presynthesized probes are applied to the substrate in a precise, predetermined volume and grid pattern, utilizing a computer-controlled robot to apply probe to the substrate in a contact-printing manner or in a non-contact manner such as ink jet or piezo-electric delivery. Probes may be covalently linked to the substrate with methods that include, but are not limited to, UV-irradiation. In another embodiment probes are linked to the substrate with heat.

Targets are nucleic acids selected from the group, including but not limited to: DNA, genomic DNA, cDNA, RNA, mRNA and may be natural or synthetic. In all embodiments,

nucleic acid molecules from human endometrial tissue are preferred. The tissue may be obtained from a subject or may be grown in culture (e.g. from a endometrial cancer cell line).

In embodiments of the invention one or more control nucleic acid molecules are attached to the substrate. Preferably, control nucleic acid molecules allow determination of factors including but not limited to: nucleic acid quality and binding characteristics; reagent quality and effectiveness; hybridization success; and analysis thresholds and success. Control nucleic acids may include but are not limited to expression products of genes such as housekeeping genes or fragments thereof.

To select a set of tumor markers, the expression data generated by, for example, microarray analysis of gene expression, preferably is analyzed to determine which genes in different groups of cancer tissues are significantly differentially expressed. In the methods disclosed herein, the significance of gene expression was determined using Permax computer software, although any standard statistical package that can discriminate significant differences in expression may be used. Permax performs permutation 2-sample t-tests on large arrays of data. For high dimensional vectors of observations, the Permax software computes t-statistics for each attribute, and assesses significance using the permutation distribution of the maximum and minimum overall attributes.

In one embodiment of the invention, expression of nucleic acid markers is used to select clinical treatment paradigms for endometrial cancer. Treatment options, as described herein, may include but are not limited to: radiotherapy, chemotherapy, adjuvant therapy, or any combination of the aforementioned methods. Aspects of treatment that may vary include, but are not limited to: dosages, timing of administration, or duration or therapy; and may or may not be combined with other treatments, which may also vary in dosage, timing, or duration. Another treatment for endometrial cancer is surgery, which can be utilized either alone or in combination with any of the aforementioned treatment methods. One of ordinary skill in the medical arts may determine an appropriate treatment paradigm based on evaluation of differential expression of sets of two or more of the nucleic acid targets ste forth as SEQ ID NOs:1-50. Cancers that express markers that are indicative of a more aggressive cancer or poor prognosis may be treated with more aggressive therapies.

Progression or regression of endometrial cancer is determined by comparison of two or more different endometrial cancer tissue samples taken at two or more different times from a subject. For example, progression or regression may be evaluated by assessments of

expression of sets of two or more of the nucleic acid targets, including but not limited to SEQ ID NOs:1-50, in an endometrial cancer tissue sample from a subject before, during, and following treatment for endometrial cancer.

In another embodiment, novel pharmacological agents useful in the treatment of endometrial cancer can be identified by assessing variations in the expression of sets of two or more endometrial cancer nucleic acid markers, from among SEQ ID NOs:1-50, prior to and after contacting endometrial cancer cells or tissues with candidate pharmacological agents for the treatment of endometrial cancer. The cells may be grown in culture (e.g. from an endometrial cancer cell line), or may be obtained from a subject, (e.g. in a clinical trial of candidate pharmaceutical agents to treat endometrial cancer). Alterations in expression of two or more sets of endometrial cancer nucleic acid markers, from among SEQ ID NOs:1-50, in endometrial cancer cells or tissues tested before and after contact with a candidate pharmacological agent to treat endometrial cancer, indicate progression, regression, or stasis of the endometrial cancer thereby indicating efficacy of candidate agents and concomitant identification of lead compounds for therapeutic use in endometrial cancer.

The invention further provides efficient methods of identifying pharmacological agents or lead compounds for agents active at the level of endometrial cancer cellular function. Generally, the screening methods involve assaying for compounds that beneficially alter endometrial cancer nucleic acid molecule expression. Such methods are adaptable to automated, high-throughput screening of compounds.

The assay mixture comprises a candidate pharmacological agent. Typically, a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a different response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e., at zero concentration of agent or at a concentration of agent below the limits of assay detection. Candidate agents encompass numerous chemical classes, although typically they are organic compounds. Preferably, the candidate pharmacological agents are small organic compounds, i.e., those having a molecular weight of more than 50 yet less than about 2500, preferably less than about 1000 and, more preferably, less than about 500. Candidate agents comprise functional chemical groups necessary for structural interactions with polypeptides and/or nucleic acids, and typically include at least an amine, carbonyl, hydroxyl, or carboxyl group, preferably at least two of the functional chemical groups and more preferably at least three of the functional chemical groups. The candidate

agents can comprise cyclic carbon or heterocyclic structure and/or aromatic or polyaromatic structures substituted with one or more of the above-identified functional groups. Candidate agents also can be biomolecules such as peptides, saccharides, fatty acids, sterols, isoprenoids, purines, pyrimidines, derivatives or structural analogs of the above, or combinations thereof and the like. Where the agent is a nucleic acid, the agent typically is a DNA or RNA molecule, although modified nucleic acids as defined herein are also contemplated.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides, synthetic organic combinatorial libraries, phage display libraries of random peptides, and the like. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are available or readily produced. Additionally, natural and synthetically produced libraries and compounds can be readily be modified through conventional chemical, physical, and biochemical means. Further, known pharmacological agents may be subjected to directed or random chemical modifications such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs of the agents.

A variety of other reagents also can be included in the mixture. These include reagents such as salts, buffers, neutral proteins (e.g., albumin), detergents, etc. which may be used to facilitate optimal protein-protein and/or protein-nucleic acid binding. Such a reagent may also reduce non-specific or background interactions of the reaction components. Other reagents that improve the efficiency of the assay such as protease, inhibitors, nuclease inhibitors, antimicrobial agents, and the like may also be used.

The mixture of the foregoing assay materials is incubated under conditions whereby, the anti-endometrial cancer candidate agent specifically binds the cellular binding target, a portion thereof or analog thereof. The order of addition of components, incubation temperature, time of incubation, and other paameters of the assay may be readily determined. Such experimentation merely involves optimization of the assay parameters, not the fundamental composition of the assay. Incubation temperatures typically are between 4°C and 40°C. Incubation times preferably are minimized to facilitate rapid, high throughput screening, and typically are between 0.1 and 10 hours.

After incubation, the presence or absence of specific binding between the anti-endometrial cancer candidate agent and one or more binding targets is detected by any convenient method available to the user. For cell-free binding type assays, a separation step is often used to separate bound from unbound components. The separation step may be accomplished in a variety of ways. Conveniently, at least one of the components is immobilized on a solid substrate, from which the unbound components may be easily separated. The solid substrate can be made of a wide variety of materials and in a wide variety of shapes, e.g., microtiter plate, microbead, dipstick, resin particle, etc. The substrate preferably is chosen to maximize signal-to-noise ratios, primarily to minimize background binding, as well as for ease of separation and cost.

Separation may be effected for example, by removing a bead or dipstick from a reservoir, emptying or diluting a reservoir such as a microtiter plate well, rinsing a bead, particle, chromotograpic column or filter with a wash solution or solvent. The separation step preferably includes multiple rinses or washes. For example, when the solid substrate is a microtiter plate, the wells may be washed several times with a washing solution, which typically includes those components of the incubation mixture that do not participate in specific bindings such as salts, buffer, detergent, non-specific protein, etc. Where the solid substrate is a magnetic bead, the beads may be washed one or more times with a washing solution and isolated using a magnet.

Detection may be effected in any convenient way for cell-based assays such as two- or three-hybrid screens. The transcript resulting from a reporter gene transcription assay of the anti-cancer agent binding to a target molecule typically encodes a directly or indirectly detectable product, e.g., β-galactosidase activity, luciferase activity, and the like. For cell-free binding assays, one of the components usually comprises, or is coupled to, a detectable label. A wide variety of labels can be used, such as those that provide direct detection (e.g., radioactivity, luminescence, optical, or electron density, etc) or indirect detection (e.g., epitope tag such as the FLAG epitope, enzyme tag such as horseseradish peroxidase, etc.). The label may be bound to an anti-cancer agent binding partner, or incorporated into the structure of the binding partner.

A variety of methods may be used to detect the label, depending on the nature of the label and other assay components. For example, the label may be detected while bound to the solid substrate or subsequent to separation from the solid substrate. Labels may be directly

detected through optical or electron density, radioactive emissions, nonradiative energy transfers, etc. or indirectly detected with antibody conjugates, strepavidin-biotin conjugates, etc. Methods for detecting the labels are well known in the art.

The invention provides endometrial cancer gene-specific binding agents, methods of identifying and making such agents, and their use in diagnosis, therapy and pharmaceutical development. For example, endometrial cancer gene-specific pharmacological agents are useful in a variety of diagnostic and therapeutic applications as described herein. In general, the specificity of an endometrial cancer gene binding to a binding agent is shown by binding equilibrium constants. Targets that are capable of selectively binding an endometrial cancer gene preferably have binding equilibrium constants of at least about 10^7 M^{-1}, more preferably at least about 10^8 M^{-1}, and most preferably at least about 10^9 M^{-1}. The wide variety of cell-based and cell-free assays may be used to demonstrate endometrial cancer gene-specific binding. Cell-based assays include one, two and three hybrid screens, assays in which endometrial cancer gene-mediated transcription is inhibited or increased, etc. Cell-free assays include endometrial cancer gene-protein binding assays, immunoassays, etc. Other assays useful for screening agents which bind endometrial cancer polypeptides include fluorescence resonance energy transfer (FRET), and electrophoretic mobility shift analysis (EMSA).

In another aspect of the invention, pre- and post-treatment alterations in expression of two or more sets of endometrial cancer nucleic acid markers including, but not limited to, SEQ ID NOs:1-50 in endometrial cancer cells or tissues may be used to assess treatment parameters including, but not limited to: dosage, method of administration, timing of administration, and combination with other treatments as described herein.

Candidate pharmacological agents may include antisense oligonucleotides that selectively bind to an endometrial cancer nucleic acid marker molecule, as identified herein, to reduce the expression of the marker molecules in endometrial cancer cells and tissues. One of ordinary skill in the art can test of the effects of a reduction of expression of endometrial cancer nucleic acid marker sequences *in vivo* or *in vitro*, to determine the efficacy of one or more antisense oligonucleotides.

As used herein, the term "antisense oligonucleotide" or "antisense" describes an oligonucleotide that is an oligoribonucleotide, oligodeoxyribonucleotide, modified oligoribonucleotide, or modified oligodeoxyribonucleotide, which hybridizes under

physiological conditions to DNA comprising a particular gene or to an mRNA transcript of that gene and, thereby, inhibits the transcription of that gene and/or the translation of that mRNA. The antisense molecules are designed so as to interfere with transcription or translation of a target gene upon hybridization with the target gene or transcript. Those skilled in the art will recognize that the exact length of the antisense oligonucleotide and its degree of complementarity with its target will depend upon the specific target selected, including the sequence of the target and the particular bases which comprise that sequence. It is preferred that the antisense oligonucleotide be constructed and arranged so as to bind selectively with the target under physiological conditions, i:e., to hybridize substantially more to the target sequence than to any other sequence in the target cell under physiological conditions.

Based upon the sequences of endometrial cancer expressed nucleic acids, or upon allelic or homologous genomic and/or cDNA sequences, one of skill in the art can easily choose and synthesize any of a number of appropriate antisense molecules for use in accordance with the present invention. In order to be sufficiently selective and potent for inhibition, such antisense oligonucleotides should comprise at least 10 and, more preferably, at least 15 consecutive bases that are complementary to the target, although in certain cases modified oligonucleotides as short as 7 bases in length have been used successfully as antisense oligonucleotides (Wagner et al., 1996). Most preferably, the antisense oligonucleotides comprise a complementary sequence of 20-30 bases. Although oligonucleotides may be chosen that are antisense to any region of the gene or mRNA transcripts, in preferred embodiments the antisense oligonucleotides correspond to N-terminal or 5' upstream sites such as translation initiation, transcription initiation, or promoter sites. In addition, 3'-untranslated regions may be targeted. Targeting to mRNA splicing sites has also been used in the art but may be less preferred if alternative mRNA splicing occurs. In addition, the antisense is targeted, preferably, to sites in which mRNA secondary structure is not expected (see, e.g., Sainio et al., 1994) and at which proteins are not expected to bind. Finally, although the listed sequences are cDNA sequences, one of ordinary skill in the art may easily derive the genomic DNA corresponding to the cDNA of an endometrial cancer expressed polypeptide. Thus, the present invention also provides for antisense oligonucleotides that are complementary to the genomic DNA corresponding to endometrial

cancer expressed nucleic acids. Similarly, the use of antisense to allelic or homologous cDNAs and genomic DNAs are enabled without undue experimentation.

In one set of embodiments, the antisense oligonucleotides of the invention may be composed of "natural" deoxyribonucleotides, ribonucleotides, or any combination thereof. That is, the 5' end of one native nucleotide and the 3' end of another native nucleotide may be covalently linked, as in natural systems, via a phosphodiester internucleoside linkage. These oligonucleotides may be prepared by art-recognized methods, which may be carried out manually or by an automated synthesizer. They also may be produced recombinantly by vectors.

In preferred embodiments, however, the antisense oligonucleotides of the invention also may include "modified" oligonucleotides. That is, the oligonucleotides may be modified in a number of ways that do not prevent them from hybridizing to their target but which enhance their stability or targeting or which otherwise enhance their therapeutic effectiveness. The term "modified oligonucleotide" as used herein describes an oligonucleotide in which (1) at least two of its nucleotides are covalently linked via a synthetic internucleoside linkage (i.e., a linkage other than a phosphodiester linkage between the 5' end of one nucleotide and the 3' end of another nucleotide) and/or (2) a chemical group not normally associated with nucleic acids has been covalently attached to the oligonucleotide. Preferred synthetic internucleoside linkages are phosphorothioates, alkylphosphonates, phosphorodithioates, phosphate esters, alkylphosphonothioates, phosphoramidates, carbamates, carbonates, phosphate triesters, acetamidates, carboxymethyl esters, and peptides.

The term "modified oligonucleotide" also encompasses oligonucleotides with a covalently modified base and/or sugar. For example, modified oligonucleotides include oligonucleotides having backbone sugars that are covalently attached to low molecular weight organic groups other than a hydroxyl group at the 3' position and other than a phosphate group at the 5' position. Thus modified oligonucleotides may include a 2'-O-alkylated ribose group. In addition, modified oligonucleotides may include sugars such as arabinose instead of ribose. The present invention, thus, contemplates pharmaceutical preparations containing modified antisense molecules that are complementary to and hybridizable with, under physiological conditions, endometrial cancer expressed nucleic acids, together with pharmaceutically acceptable carriers.

Antisense oligonucleotides may be administered as part of a pharmaceutical composition. Such a pharmaceutical composition may include the antisense oligonucleotides in combination with any standard physiologically and/or pharmaceutically acceptable carriers which are known in the art. The compositions should be sterile and contain a therapeutically effective amount of the antisense oligonucleotides in a unit of weight or volume suitable for administration to a patient. The term "pharmaceutically acceptable" means a non-toxic material that does not interfere with the effectiveness of the biological activity of the active ingredients. The term "physiologically acceptable" refers to a non-toxic material that is compatible with a biological system such as a cell, cell culture, tissue, or organism. The characteristics of the carrier will depend on the route of administration. Physiologically and pharmaceutically acceptable carriers include diluents, fillers, salts, buffers, stabilizers, solubilizers, and other materials, which are well known in the art.

Expression of endometrial cancer nucleic acid molecules can also be determined using protein measurement methods to determine expression of SEQ ID NOs:1-50, e.g., by determining the expression of polypeptides encoded by SEQ ID NOs:1-50 (SEQ ID NOs: 51-100). Preferred methods of specifically and quantitatively measuring proteins include, but are not limited to: mass spectroscopy-based methods such as surface enhanced laser desorption ionization (SELDI; e.g., Ciphergen ProteinChip System), non-mass spectroscopy-based methods, and immunohistochemistry-based methods such as 2-dimensional gel electrophoresis.

SELDI methodology may, through procedures known to those of ordinary skill in the art, be used to vaporize microscopic amounts of tumor protein and to create a "fingerprint" of individual proteins, thereby allowing simultaneous measurement of the abundance of many proteins in a single sample. Preferably SELDI-based assays may be utilized to classify endometrial cancer tumors. Such assays preferably include, but are not limited to the following examples. Gene products discovered by RNA microarrays may be selectively measured by specific (antibody mediated) capture to the SELDI protein disc (e.g., selective SELDI). Gene products discovered by protein screening (e.g., with 2-D gels), may be resolved by "total protein SELDI" optimized to visualize those particular markers of interest from among SEQ ID NOs:1-50. Predictive models of tumor classification from SELDI measurement of multiple markers from among SEQ ID NOs:1-50 may be utilized for the SELDI strategies. In an additional embodiment a set of endometrioid endometrial

adenocarcinoma tissues may be preferably utilized to determine the risk classification of endometrial cancer based on SELDI results.

The invention also involves agents such as polypeptides that bind to endometrial cancer-associated polypeptides, i.e., SEQ ID NOs:51-100. Such binding agents can be used, for example, in screening assays to detect the presence or absence of endometrial cancer-associated polypeptides and complexes of endometrial cancer-associated polypeptides and their binding partners and in purification protocols to isolate endometrial cancer-associated polypeptides and complexes of endometrial cancer-associated polypeptides and their binding partners. Such agents also may be used to inhibit the native activity of the endometrial cancer-associated polypeptides, for example, by binding to such polypeptides.

The invention, therefore, embraces peptide binding agents which, for example, can be antibodies or fragments of antibodies having the ability to selectively bind to endometrial cancer-associated polypeptides. Antibodies include polyclonal and monoclonal antibodies, prepared according to conventional methodology.

Significantly, as is well-known in the art, only a small portion of an antibody molecule, the paratope, is involved in the binding of the antibody to its epitope (see, in general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology Wiley & Sons, Inc., New York; Roitt, I. (1991) Essential Immunology, 7th Ed., Blackwell Scientific Publications, Oxford). The pFc' and Fc regions, for example, are effectors of the complement cascade but are not involved in antigen binding. An antibody from which the pFc' region has been enzymatically cleaved, or which has been produced without the pFc' region, designated an $F(ab')_2$ fragment, retains both of the antigen binding sites of an intact antibody. Similarly, an antibody from which the Fc region has been enzymatically cleaved, or which has been produced without the Fc region, designated an Fab fragment, retains one of the antigen binding sites of an intact antibody molecule. Proceeding further, Fab fragments consist of a covalently bound antibody light chain and a portion of the antibody heavy chain denoted Fd. The Fd fragments are the major determinant of antibody specificity (a single Fd fragment may be associated with up to ten different light chains without altering antibody specificity) and Fd fragments retain epitope-binding ability in isolation.

Within the antigen-binding portion of an antibody, as is well-known in the art, there are complementarity determining regions (CDRs), which directly interact with the epitope of the antigen, and framework regions (FRs), which maintain the tertiary structure of the

paratope (see, in general, Clark, 1986; Roitt, 1991). In both the heavy chain Fd fragment and the light chain of IgG immunoglobulins, there are four framework regions (FR1 through FR4) separated respectively by three complementarity determining regions (CDR1 through CDR3). The CDRs, and in particular the CDR3 regions, and more particularly the heavy chain CDR3, are largely responsible for antibody specificity.

It is now well-established in the art that the non-CDR regions of a mammalian antibody may be replaced with similar regions of conspecific or heterospecific antibodies while retaining the epitopic specificity of the original antibody. This is most clearly manifested in the development and use of "humanized" antibodies in which non-human CDRs are covalently joined to human FR and/or Fc/pFc' regions to produce a functional antibody. See, e.g., U.S. patents 4,816,567, 5,225,539, 5,585,089, 5,693,762 and 5,859,205.

Fully human monoclonal antibodies also can be prepared by immunizing mice transgenic for large portions of human immunoglobulin heavy and light chain loci. Following immunization of these mice (e.g., XenoMouse (Abgenix), HuMAb mice (Medarex/GenPharm)), monoclonal antibodies can be prepared according to standard hybridoma technology. These monoclonal antibodies will have human immunoglobulin amino acid sequences and therefore will not provoke human anti-mouse antibody (HAMA) responses when administered to humans.

Thus, as will be apparent to one of ordinary skill in the art, the present invention also provides for $F(ab')_2$, Fab, Fv and Fd fragments; chimeric antibodies in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric $F(ab')_2$ fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; chimeric Fab fragment antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; and chimeric Fd fragment antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced by homologous human or non-human sequences. The present invention also includes so-called single chain antibodies.

Thus, the invention involves polypeptides of numerous size and type that bind specifically to polypeptides selected from SEQ ID NOs:51-100, and complexes of both endometrial cancer-associated polypeptides and their binding partners. These polypeptides may be derived also from sources other than antibody technology. For example, such

polypeptide binding agents can be provided by degenerate peptide libraries which can be readily prepared in solution, in immobilized form or as phage display libraries. Combinatorial libraries also can be synthesized of peptides containing one or more amino acids. Libraries further can be synthesized of peptoids and non-peptide synthetic moieties.

Phage display can be particularly effective in identifying binding peptides useful according to the invention. Briefly, one prepares a phage library (using e.g. m13, fd, or lambda phage), displaying inserts from 4 to about 80 amino acid residues using conventional procedures. The inserts may represent, for example, a completely degenerate or biased array. One then can select phage-bearing inserts which bind to the endometrial cancer-associated polypeptide. This process can be repeated through several cycles of reselection of phage that bind to the endometrial cancer-associated polypeptide. Repeated rounds lead to enrichment of phage bearing particular sequences. DNA sequence analysis can be conducted to identify the sequences of the expressed polypeptides. The minimal linear portion of the sequence that binds to the endometrial cancer-associated polypeptide can be determined. One can repeat the procedure using a biased library containing inserts containing part or all of the minimal linear portion plus one or more additional degenerate residues upstream or downstream thereof. Yeast two-hybrid screening methods also may be used to identify polypeptides that bind to the endometrial cancer-associated polypeptides.

Thus, the endometrial cancer-associated polypeptides of the invention, including fragments thereof, can be used to screen peptide libraries, including phage display libraries, to identify and select peptide binding partners of the endometrial cancer-associated polypeptides of the invention. Such molecules can be used, as described, for screening assays, for purification protocols, for interfering directly with the functioning of endometrial cancer-associated polypeptides and for other purposes that will be apparent to those of ordinary skill in the art. For example, isolated endometrial cancer-associated polypeptides can be attached to a substrate (e.g., chromatographic media, such as polystyrene beads, a filter, or an array substrate), and then a solution suspected of containing the binding partner may be applied to the substrate. If a binding partner that can interact with endometrial cancer-associated polypeptides is present in the solution, then it will bind to the substrate-endometrial cancer-associated polypeptide. The binding partner then may be isolated.

As detailed herein, the foregoing antibodies and other binding molecules may be used for example, to identify tissues expressing protein or to purify protein. Antibodies also may

be coupled to specific diagnostic labeling agents for imaging of cells and tissues that express endometrial cancer-associated polypeptides or to therapeutically useful agents according to standard coupling procedures. Diagnostic agents include, but are not limited to, barium sulfate, iocetamic acid, iopanoic acid, ipodate calcium, diatrizoate sodium, diatrizoate meglumine, metrizamide, tyropanoate sodium and radiodiagnostics including positron emitters such as fluorine-18 and carbon-11, gamma emitters such as iodine-123, technitium-99m, iodine-131 and indium-111, nuclides for nuclear magnetic resonance such as fluorine and gadolinium.

The invention further includes protein microarrays for analyzing expression of endometrial cancer-associated peptides selected from SEQ ID NOs:51-100. In this aspect of the invention, standard techniques of microarray technology are utilized to assess expression of the endometrial cancer-associated polypeptides and/or identify biological constituents that bind such polypeptides. The constituents of biological samples include antibodies, lymphocytes (particularly T lymphocytes), and the like. Protein microarray technology, which is also known by other names including: protein chip technology and solid-phase protein array technology, is well known to those of ordinary skill in the art and is based on, but not limited to, obtaining an array of identified peptides or proteins on a fixed substrate, binding target molecules or biological constituents to the peptides, and evaluating such binding. See, e.g., G. MacBeath and S.L. Schreiber, "Printing Proteins as Microarrays for High-Throughput Function Determination," *Science* 289(5485):1760-1763, 2000.

Preferably antibodies or antigen binding fragments thereof that specifically bind polypeptides selected from the group consisting of SEQ ID NOs:51-100 are attached to the microarray substrate in accordance with standard attachment methods known in the art. These arrays can be used to quantify the expression of the polypeptides identified herein.

In some embodiments of the invention, one or more control peptide or protein molecules are attached to the substrate. Preferably, control peptide or protein molecules allow determination of factors such as peptide or protein quality and binding characteristics, reagent quality and effectiveness, hybridization success, and analysis thresholds and success.

The use of such methods to determine expression of endometrial cancer nucleic acids from among SEQ ID NOs:1-50 and/or proteins from among SEQ ID Nos:51-100 can be done with routine methods known to those of ordinary skill in the art and the expression

determined by protein measurement methods may be used as a prognostic method for selecting treatment strategies for endometrial cancer patients.

Examples

To establish a prognostic tool for designing endometrial cancer treatment regimens, expression patterns in primary endometrialcancer specimens were assessed and correlated with clinical outcome.

Tissue processing:

RNA isolated from normal cycling (proliferative, n=2; secretory, n=2) and neoplastic (endometrioid adenocarcinoma, n=10) human endometrial specimens was reverse transcribed and resultant cDNA used for in *vitro* transcriptional synthesis of fluorescently labeled nucleic acid probes according to manufacturer's instructions. Each resultant tissue-derived probe was then separately hybridized to an Affymetrix HuFL human expression array and hybridization images analyzed with Affymetrix software to generate a data matrix of named probes by quantitative expression level in each tissue.

Data Normalization:

Average differences for each sample were rescaled to sum to 3,000,000 over all genes. Then the average differences with an Affymetrix call of Absent or Marginal were set to 20, and average differences with a call of Present but with less than 20 were also set to 20. This resulted in a dataset truncated on the left tail at a value of 20, in which only genes determined to be "present" by the Affymetrix call were included as positive expression values.

Permax Test:

Standard pooled variance t-statistics comparing the 4 normal samples to the 10 tumor samples were computed separately for each gene from their log values. Log values were used because it is natural to think of differences between tissue types as a multiplicative effect or ratio increase/decrease. Only genes with at least 2 values > 20 were considered (3665 genes), since the t statistic is undefined for genes with all values = 20, and the statistic is either 1.69 or -.62 with only one value not equal to 20, regardless of the value.

The permutation distribution was used to assess the significance of t-statistics calculated for each gene in the dataset (Permax test). The customized program written in S-plus language to calculate Permax is a data analysis software tool for testing the significance of gene expression. It has been presented by Mutter, et al., 8th International Workshop on Chromosomes in Solid Tumors, Tucson, AZ, 2000; and is available online[2] at biowww.dfci.harvard.edu/~gray/permax.html and from Robert J. Gray, Department of Biostatistical Science, Dana-Farber Cancer Institute, 44 Binney Street Boston, MA 02115. Permax details enclosed therein are incorporated by reference herein.. In this approach all 1001 possible ways of dividing the 14 samples into two groups of sizes 4 and 10 were considered. For each of these, the t-statistics were computed for each gene. With unequal group sizes, these distributions are not symmetric, so the significance was assessed separately in each direction. To control the overall error rate, the distributions of the maximum and minimum t-statistics over the genes were used. That is, for each gene, the p-value in the direction with expression higher (lower) in normals is the proportion of permutations with the minimum (maximum) t statistic over all genes less than (greater than) or equal to the observed value for the particular gene. A test declaring as significant any genes with say $p<.50$ then guarantees that the chance of any false positives being selected is <50%.

The t statistics have a tendency to preferentially select genes with very small variances within a group. Because of this it may be appropriate to also require minimum criteria for differences between the group means. After determining the most significant genes from the t statistics, those genes with absolute differences between means <100, and ratios of means <3 were identified.

Table 1 is a spreadsheet identifying 50 genes which discriminate normal cycling from malignant endometrium.

Table 1

SEQ ID NO	GeneCode	Permax GPT	Fold GPT	Delta GPT	ChrBand	NLX GPT	TX GPT	AffyProbe Set	LocusLink	GenBank	ABREV	Title (from Unigene)
1	x6235	0.042	8.9	157.3	17q21	177	20	D88213_at	314	D88213	AOC2	amine oxidase, copper containing 2 (retina-specific)
2	x4535	0.2218	11.6	344.8	19q13.1	377	32	HG162-HT3165_at	558	M76125	AXL	AXL receptor tyrosine kinase
3	x2035	0.2727	45.9	898.1	11p15.5	20	918	M91083_at	8045	M91083	C11ORF13	chromosome 11 open reading frame 13
4	x3265	0.468	10.1	1590.5	12p13	1766	175	D13639_at	894	D13639	CCND2	cyclin D2
5	x3120	0.5	8.8	446.4	16q22.1	504	57	D21255_at	1009	D21255	CDH11	cadherin 11 (OB-cadherin, osteoblast)
6	x6580	0.2587	8.9	255.1	1p21	287	32	J04177_at	1301	J04177	COL11A1	collagen, type XI, alpha 1
7	x2140	0.1938	13.3	412.2	8q23	446	33	Y11710_rna1_at	7373	Y11710	COL14A1	collagen, type XIV, alpha 1; undulin
8	x1629	0.2038	8.9	158.8	2p21	179	20	U03688_at	1545	U03688	CYP1B1	cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1 (glaucoma 3, primary infantile)
9	x3108	0.028	13.9	258.1	17p13.1	278	20	U83192_at	1742	U83192	DLG4	discs, large (Drosophila) homolog 4
10	x3342	0.426	6.5	3499.8	5q34	4140	640	X68277_at	1843	X68277	DUSP1	dual specificity phosphatase 1
11	x4985	0.2448	4.4	113.1	8	33	146	U15642_s_at	1875	U15642	E2F5	E2F transcription factor 5, p130-binding
12	x671	0.446	11.2	597.5	4	656	58	D11151_at	1909	D11151	EDNRA	endothelin receptor type A
13	x2341	0.2448	5.5	1078.6	8p21.1	242	1321	HG4535-HT4940_s_at	2039	U28389	EPB49	erythrocyte membrane protein band 4.9 (dematin)
14	x2797	0.0959	25.4	489.0	16p13.3-p13.11	20	509	L76568_xpt3_f_at	2072	L76568	ERCC4	excision repair cross-complementing rodent repair deficiency, complementation group 4
15	x6244	0.3057	3.1	750.5	13q14.1-q14.2	1103	353	M13450_at	2098	M13450	ESD	esterase D/formylglutathione hydrolase
16	x2404	0.2128	8.7	245.1	Xq22	277	32	X97249_at	2491	X97249	FSHPRH1	FSH primary response (LRPR1, rat) homolog 1
17	x4516	0.3247	39.1	761.3	3p21.3	20	781	U49082_at	10991	U49082	G17	G17 transporter protein
18	x4495	0.2218	55.8	3521.2	2p12-q11	3585	64	M85276_at	10578	M85276	GNLY	granulysin
19	x1222	0.014	16.5	310.3	2q14-q21	330	20	M36284_s_at	2995	M36284	GYPC	glycophorin C (Gerbich blood group)
20	x2590	0.1359	7.5	129.1	15q22	149	20	U50078_at	8925	U50078	HERC1	hect (homologous to the E6-AP (UBE3A) carboxyl terminus) domain

												and RCC1 (CHC1)-like domain (RLD) 1
21	x881	0.0599	10.3	185.2	2	205	20	U44111_at	3176	U44111	HNMT	histamine N-methyltransferase
22	x5023	0.2388	11.2	395.1	2q37.1-q36.3	434	39	X77307_at	3357	X77307	HTR2B	5-hydroxytryptamine (serotonin) receptor 2B
23	x2719	0.2108	7.9	206.0	9p22	236	30	J00212_f_at	3452	V00540	IFNA21	interferon, alpha 21
24	x5442	0.1618	42.3	1454.6	12q22-q23	1490	35	X57025_at	3479	X57025	IGF1	insulin-like growth factor 1 (somatomedin C)
25	x5452	0.5	6.2	420.4	19p13.1	501	81	U61263_at	10994	U61263	ILVBL	ilvB (bacterial acetolactate synthase)-like
26	x6197	0.1808	10.4	188.0	4q34.1-q35.1	208	20	X15949_at	3660	X15949	IRF2	interferon regulatory factor 2
27	x3700	0.3447	7.2	124.9	3q21-q25	145	20	D13626_at	9934	D13626	KIAA0001	KIAA0001 gene product
28	x1553	0.1728	3.7	3035.7	12q13	1115	4151	X12876_s_at	3875	X12876	KRT18	keratin 18
29	x5912	0.0669	10.2	4816.8	12q13	521	5338	X74929_s_at	3856	X74929	KRT8	keratin 8
30	x197	0.035	43.7	853.4	16q22.1	873	20	M12625_at	3931	M12625	LCAT	lecithin-cholesterol acyltransferase
31	x723	0.2478	5.8	7378.3	22q13.1	8915	1537	J04456_at	3956	J04456	LGALS1	lectin, galactoside-binding, soluble, 1 (galectin 1)
32	x1271	0.1299	3.8	1609.5	4q	583	2192	M93036_at	4072	M93036	M4S1	membrane component, chromosomal 4, surface marker (35kD glycoprotein)
33	x6752	0.431	4.0	613.2	14	818	205	Z24725_at	10979	Z24725	MIG2	mitogen inducible 2
34	x1469	0.2038	18.2	797.4	14q11-q12	844	46	Z48481_at	4323	Z48481	MMP14	matrix metalloproteinase 14 (membrane-inserted)
35	x879	0.2458	22.3	1149.7	1q43	1204	54	M30269_at	4811	M30269	NID	nidogen (enactin)
36	x1397	0.3946	11.6	724.8	5q14	793	68	HG3510-HT3704_at	7025	X12795	NR2F1	nuclear receptor subfamily 2, group F, member 1
37	x2831	0.2987	6.0	204.3	10q21.3-q23.1	41	245	M24486_s_at	5033	M24486	P4HA1	procollagen-proline, 2-oxoglutarate 4-dioxygenase (proline 4-hydroxylase), alpha polypeptide 1
38	x2670	0.2478	158.2	4966.1	9q34	4998	32	HG721-HT4827_s_at	5047	J04129	PAEP(alt2)	progestagen-associated endometrial protein (placental protein 14, pregnancy-associated endometrial alpha-2-globulin, alpha uterine protein): Alternate Splice 2
39	x5757	0.2038	12.3	3977.7	7q22	4331	353	L33799_at	5118	L33799	PCOLCE	procollagen C-endopeptidase enhancer

40	x6701	0.3077	17.8	1101.4	4q11-q13	1167	65	M21574_at	5156	M21574	PDGFRA	platelet-derived growth factor receptor, alpha polypeptide
41	x6741	0.2068	5.1	120.6	Xq21-q27	150	29	D00860_at	5631	D00860	PRPS1	phosphoribosyl pyrophosphate synthetase 1
42	x1195	0.1099	10.2	183.7	Xp22.3-p22.2	204	20	Y00971_at	5634	Y00971	PRPS2	phosphoribosyl pyrophosphate synthetase 2
43	x5284	0.0789	11.6	212.7	4p15.31	233	20	M16447_at	5860	M16447	QDPR	quinoid dihydropteridine reductase
44	x320	0.4076	5.6	167.7	1p31-p22	204	37	X98001_at	5876	X98001	RABGGTB	Rab geranylgeranyltransferase, beta subunit
45	x6986	0.3417	13.9	628.3	10q11.1	677	49	L36033_at	6387	L36033	SDF1	stromal cell-derived factor 1
46	x1047	0.1988	3.7	297.9	5q31	408	110	Z11793_at	6414	Z11793	SEPP1	selenoprotein P, plasma, 1
47	x4685	0.2218	8.3	215.8	Xq28	245	30	X92396_at	6845	X92396	SYBL1	synaptobrevin-like 1
48	x5624	0.2228	3.1	666.3	15q13	988	321	L14837_at	7082	L14837	TJP1	tight junction protein 1 (zona occludens 1)
49	x4880	0.038	13.0	239.8	11p13	260	20	X69950_s_at	51352	X69950	WIT-1	Wilms tumor associated protein
50	x860	0.1508	13.9	323.7	7q22-q32	349	25	X98260_at	27000	X98260	ZRF1	zuotin related factor 1

Key:

SEQ ID NO	Sequence identifier number
GeneCode	Internal lab unique identifier, numbers preceded by an "x"
PermaxGPT	Permax value using GPT datastate
FoldGPT	Ratio of NLXGPT to TXGPT, inverted if needed to yield value >1
DeltaGPT	Arithmetic difference of NLXGPT and TXGPT , absolute value
ChrBand	Karyotypic locus of gene
NLXGPT	Mean expression in GPT units of 4 normal endometria
TXGPT	Mean expression in GPT units of 10 endometrioid endometrial adenocarcinomas
AffyProbeSet	Affymetrix probe identifier in HuFL human expression array chip
LocusLink	Locuslink ID number, when available.
GenBank	The GenBank entry for sequence used by Affymetrix to design probes
Abrev	When in full caps, this is the Locuslink recommended nomenclature.
Title	Text description of gene. Usually Locuslink label

References

1. Mutter, G. L., Baak, J. P. A., Cai, T., Fitzgerald, J., Gray, R., Gentleman, R., Gullans, S., Ibrahim, J., Neuberg, D., and Wilcox, M. Altered Gene Expression in Endometrioid Endometrial Adenocarcinomas Analyzed by High Density Microarrays. 8th International Workshop on Chromosomes in Solid Tumors (Tucson,AZ) . 2000. Ref Type: Abstract
2. Gray, R. Permax [On Line]. Available: http://biowww.dfci.harvard.edu/~gray/permax.html . 2000.
3. National Center for Biotechnology Information (USA): Unigene [Online]. Available: http://www ncbi nlm nih gov/UniGene/ 2000.
4. Wheeler DL, Chappey C, Lash AE, Leipe DD, Madden TL, Schuler GD, Tatusova TA, Rapp BA: Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2000, 28:10-14.
5. National Center for Biotechnology Information (USA): Locuslink [Online]. Available: http://www ncbi nlm nih gov/LocusLink/ 2000.
6. Pruitt KD, Katz KS, Sicotte H, Maglott DR: Introducing RefSeq and LocusLink: curated human genome resources at the NCBI. Trends Genet 2000, 16:44-47.
7. Harrison's Principles of Internal Medicine, 14/e, (1998) McGraw-Hill Companies, New York.
8. The Chipping Forecast (1999) Nature Genetics, 21(1):1-60.
9. Gwynne, P., and Page, G., (1999) Microarray Analysis: the next revolution in Molecular Biology, Science eMarketplace, Science, August 6. (sciencemag.org/feature/e-market/benchtop/micro.shl)
10. *Molecular Cloning: A Laboratory Manual*, (1989) J. Sambrook, et al., eds., Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
11. *Current Protocols in Molecular Biology*, (1999) F.M. Ausubel, et al., eds., John Wiley & Sons, Inc., New York.
12. Wagner et al., Nature Biotechnol. (1996) 14:840-844.
13. Sainio, K., Saarma, M., Nonclercq, D., Paulin, L., and Sariola, H. (1994) Antisense inhibition of low-affinity nerve growth factor receptor in kidney cultures: power and pitfalls.Cell Mol. Neurobiol. 14(5):439-457.

The present invention is not limited in scope by the examples provided, since the examples are intended as illustrations of various aspects of the invention and other functionally equivalent embodiments are within the scope of the invention. Various modifications of the invention in addition to those shown are described herein will become apparent to those skilled in the art for the foregoing description and fall within the scope of the appended claims. The advantages and objects of the invention are not necessarily encompassed by each embodiment of the invention.

All references, patents, and patent publications that are recited in this application are incorporated in their entirety herein by reference.

I claim: